

## An iron (II) dependent oxygenase performs the last missing step of plant lysine catabolism

Mitchell G. Thompson,<sup>1,2,3\*</sup> Jacquelyn M. Blake-Hedges,<sup>1,2,4\*</sup> Jose Henrique Pereira,<sup>1,5\*</sup> John A. Hangasky,<sup>4</sup> Michael S. Belcher,<sup>1,2,3</sup> William M. Moore,<sup>1,2,3</sup> Jesus F. Barajas,<sup>1,2,6</sup> Pablo Cruz-Morales,<sup>1,2</sup>, Lorenzo J. Washington,<sup>1,2,3</sup> Robert W. Haushalter,<sup>1,2</sup> Christopher B. Eiben<sup>1,2,11</sup>, Yuzhong Liu,<sup>1,2</sup> Will Skyrud,<sup>4</sup> Veronica T. Benites,<sup>1,2</sup> Tyler P. Barnum,<sup>3</sup> Edward E. K. Baidoo,<sup>1,2</sup> Henrik V. Scheller,<sup>1,2,3</sup> Michael A. Marletta,<sup>4,7</sup> Patrick M. Shih,<sup>1,2,8,9,10</sup> Paul D. Adams,<sup>1,5,11</sup> Jay D. Keasling<sup>1,2,11,12,13,14,#</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, CA, USA

<sup>2</sup>Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>3</sup>Department of Plant and Microbial Biology, University of California-Berkeley, Berkeley, CA, USA

<sup>4</sup>Department of Chemistry, University of California-Berkeley, Berkeley, CA, USA

<sup>5</sup>Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>6</sup>Department of Energy Agile BioFoundry, Emeryville, CA, USA

<sup>7</sup>Department of Molecular and Cellular Biology, University of California-Berkeley, Berkeley, CA, USA

<sup>8</sup>Department of Plant Biology, University of California-Davis, Davis, CA, USA

<sup>9</sup>Genome Center, University of California-Davis, Davis, CA, USA

<sup>10</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>11</sup>Department of Bioengineering, University of California-Berkeley, Berkeley, CA 94720, USA

<sup>12</sup>Department of Chemical and Biomolecular Engineering, University of California-Berkeley, Berkeley, CA, USA

<sup>13</sup>The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark

<sup>14</sup>Center for Synthetic Biochemistry, Shenzhen Institutes for Advanced Technologies, Shenzhen, China

\*Authors contributed equally

# Corresponding author jdkeasling@lbl.gov

## Abstract

Due to low abundance in many staple food crops, the essential amino acid lysine must be produced industrially to meet global food supply needs. Despite intensive study, manipulation, and agricultural importance, the steps of plant lysine catabolism beyond the 2-oxoadipate (2OA) intermediate remain undescribed. Recently we described a missing step in the D-lysine catabolic pathway of the bacterium *Pseudomonas putida* in which 2OA is converted to D-2-hydroxyglutarate (D2HG) via hydroxyglutarate synthase (HglS), an enzyme belonging to the previously uncharacterized DUF1338 protein family. Here we solve the structure of HglS to 1.1 Å resolution in the substrate-free form and in complex with 2OA. Structural similarity to hydroxymandelate synthase suggested a successive decarboxylation and intramolecular hydroxylation mechanism forming 2HG in a Fe(II)- and O<sub>2</sub>-dependent manner, which is validated experimentally. 2OA specificity was mediated by a single arginine (R74), highly conserved across nearly all DUF1338 family proteins, including in 76% of plant enzymes. In *Arabidopsis thaliana*, a DUF1338 homolog is coexpressed with known lysine catabolism enzymes, and mutants show significant germination rate defects consistent with disrupted lysine catabolism. Structural and biochemical analysis of the *Oryza sativa* homolog FLO7 revealed identical activity to HglS despite low sequence identity. Our results suggest that nearly all DUF1338 containing enzymes likely catalyze the same biochemical reaction, exerting the same physiological function across bacteria and eukaryotes.

## Significance

**To meet human demands, millions of tons of lysine are produced by bacterial fermentation annually due to its low abundance in staple crops. Here, we show the last missing step of lysine catabolism in nearly all plant endosperms is likely catalyzed by an**

**iron-dependant DUF1338-containing enzyme homologous to the bacterial hydroxyglutarate synthase. Structural and bioninformatic analyses of DUF1338-containing enzymes showed high conservation of critical catalytic and specificity-conferring residues across multiple domains of life despite low sequence identity. These results suggest that the DUF1338 family evolved a specific physiological function within lysine catabolism across multiple domains of life.**

## **Introduction**

Lysine is an essential amino acid, and due to its low abundance in cereals and legumes it is produced on a scale of one million tons a year to supplement food supply needs.<sup>1-3</sup> To thwart malnutrition in the developing world, significant work has been done to engineer rice, maize and other plants to produce greater quantities of lysine.<sup>3,4</sup> Increasing lysine levels in cereal grains requires overexpression of lysine-producing enzymes and concurrent disruption of lysine catabolism.<sup>3</sup> Thus, mutants such as *opaque2* in maize have received considerable attention for their ability to accumulate lysine within their endosperm.<sup>5,6</sup> However, despite worldwide importance, the full plant lysine catabolism pathway remains unknown, with no consensus in the steps beyond 2-oxoadipate (2OA) formation.<sup>7</sup>

Recently we described a novel D-lysine catabolic route in the bacterium *Pseudomonas putida* which also contains a 2OA intermediate, similar to plant L-lysine catabolism.<sup>8</sup> In the *P. putida* pathway, 2OA is converted to 2-oxoglutarate (2OG) via three enzymes, one of which catalyzes a unique decarboxylation-hydroxylation step. In this step, 2OA is converted to D-2-hydroxyglutarate (2HG) in a reaction catalyzed by the Fe(II)-dependent DUF1338 family enzyme HglS (Figure S1).<sup>8</sup> Homologs of this enzyme are broadly distributed across multiple

domains of life, including nearly every sequenced plant genome.<sup>8,9</sup> However, only one study describing DUF1338 enzymes in plants has been reported. In *Oryza sativa* *FLO7* (a DUF1338 family member) mutants, abnormal starch formation was observed in the endosperm similar to maize *opaque2* mutants which are known to accumulate high levels of lysine.<sup>9</sup> The widespread DUF1338 family abundance in plants and the *FLO7* and *opaque2* phenotypes encouraged us to further investigate whether enzymes throughout the family displayed similar activity to Hg1S.

Conversion of 2OA directly to D2HG requires two discrete chemical steps, a decarboxylation and hydroxylation, making the chemical mechanism of the enzyme puzzling.<sup>8</sup> Here, we leverage structural and biochemical analyses to postulate a chemical mechanism for Hg1S and characterize its substrate specificity. We show that critical residues involved in catalysis are highly conserved across nearly all DUF1338 family proteins. We further show that despite very low sequence identity, plant homologs also catalyze the conversion of 2OA to 2HG and adopt the same structural fold as the bacterial enzyme, suggesting that DUF1338 family proteins catalyze the last unknown step of plant lysine catabolism.

## Results

### Structural analysis reveals the catalytic and specificity mechanisms of Hg1S

To better understand the unusual Hg1S reaction, we obtained crystal structures of the enzyme both with and without a bound substrate. Initially, Hg1S was crystallized without substrate, and the Hg1S structure was solved at 1.1 Å resolution (Figure 1A). The enzyme possesses a central β-sheet motif resembling a partially-closed β-barrel consisting of 7 β-sheets which is conserved in the three other DUF1338 structures deposited in the Protein Data Bank (PDB IDs 3LHO, 3IUZ and 2RJB).<sup>10-12</sup> Further analysis also revealed the presence of a metal cofactor bound within a conserved metal cofactor-binding motif—consisting of the residues His

70, His 226, and Glu 294—common to the DUF1338 structures. However, none of the available DUF1338 structures have been biochemically characterized or solved in complex with a substrate, and consequently no chemical reaction mechanism for the family has been proposed.

Therefore, we performed a search for characterized proteins with similar structures using the Vector Alignment Search Tool (VAST).<sup>13</sup> The top VAST hits were the three reported DUF1338 structures, followed by the hydroxymandelate synthase (HMS) and 4-hydroxyphenyl pyruvate dioxygenase (HPPD) structures (Table S1).<sup>14,15</sup> HgLS and HMS structure comparison revealed the two enzymes share a similar central  $\beta$ -sheet fold common to the DUF1338 structures (Figure S2).<sup>14,16</sup> More importantly, the  $\beta$ -sheet domain of HMS contains the enzyme active site, two histidines and a glutamate that bind the metal cofactor in nearly the same orientation as in HgLS. The similarity of the HgLS, HMS, and HPPD folds and active sites suggested that HgLS is likely an additional member of the vicinal oxygen chelate (VOC) enzyme superfamily and could act via a similar mechanism to HMS and HPPD.<sup>17,18</sup>

To determine metal cofactor identity within the HgLS structure, we conducted a fluorescent scan of a HgLS crystal. We detected a photon emission near 7475 eV, the K-alpha emission of nickel (Figure S3). In addition, a less intense peak corresponding to the K-alpha emission of iron was also observed. We therefore assigned the metal cofactor as Ni(II) with a small percent Fe(II) occupancy. While the metal cofactors present in the HMS and HgLS structures differ, we previously determined that HgLS utilizes Fe(II), not Ni(II), for catalysis. The nickel bound to HgLS is likely derived from the nickel affinity chromatography protein purification.

The HgLS domain architecture suggests the enzyme belongs to the VOC superfamily, mandating a chemical mechanism employing the bidentate coordination of vicinal oxygen atoms

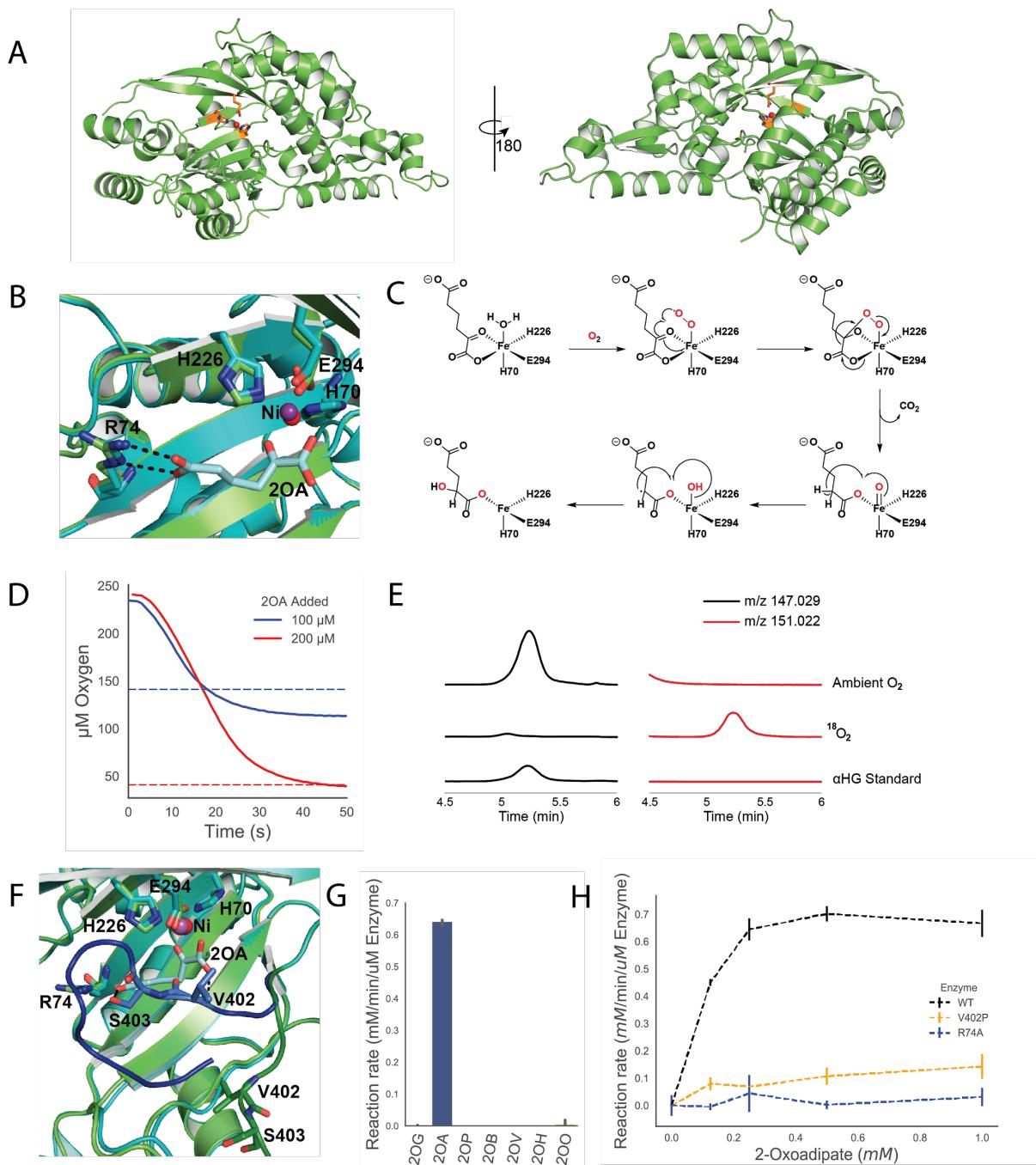
to a divalent metal center. Given the established HMS substrate binding mode, we hypothesized that the metal in HgIS would bind the vicinal oxygen atoms of the  $\alpha$ -keto group of 2OA.<sup>14</sup> We therefore soaked HgIS crystals with 2OA and solved the structure of the resulting complex. The noncatalytic active site nickel prevented enzyme turnover, yielding a substrate-bound structure and no observed product density. As hypothesized, the substrate carboxylate and  $\alpha$ -keto group oxygens are coordinated to the metal (Figure 1B), classifying HgIS as a VOC superfamily member.

We had previously noted the similarity between the set of reactions catalyzed by HgIS and HMS.<sup>8,14,16</sup> Both enzymes perform a decarboxylation-hydroxylation reaction on a  $\alpha$ -ketoacid substrate. HPPD catalyzes a similar overall reaction with the hydroxylation occurring at a different position. The structural and biochemical similarity of HgIS to HMS and HPPD led us to propose a similar chemical mechanism for HgIS (Figure 1C). More specifically, the catalytic cycle begins when the substrate carboxylate and  $\alpha$ -keto group oxygens and molecular oxygen bind to the three open Fe(II) coordination sites. A radical rearrangement results in a decarboxylation and the formation of a Fe(IV)-oxo and an  $\alpha$ -radical species. Finally, continued radical rearrangement produces the hydroxylated product 2-hydroxyglutarate. If HgIS follows this mechanism, it should consume 1 mole of O<sub>2</sub> per mole of 2OA, and two oxygens in the product should derive from molecular O<sub>2</sub>. To support this mechanism, we determined the stoichiometry of 2OA to O<sub>2</sub> consumption using a dissolved oxygen probe. In the presence of 100 or 200  $\mu$ M 2OA, the enzyme consumed approximately an equimolar concentration of O<sub>2</sub> (Figure 1D). No feedback inhibition was observed with 1 mM D-2HG, or L-2HG (Figure S4A), and kinetic parameters determined by monitoring oxygen consumption were similar to those we previously reported using a different enzyme assay (Figure S4B).<sup>8</sup> We next tested our hypothesis

that two oxygen atoms in the 2HG product derive from molecular O<sub>2</sub>. To determine the source of oxygen atoms, additional enzyme assays were performed under an <sup>18</sup>O<sub>2</sub> atmosphere. High resolution LC-MS product analysis revealed a species exhibiting a m/z of 151.022 that co-eluted with a 2HG standard (m/z 147.029), corresponding to the expected m/z of 2HG containing two <sup>18</sup>O-labeled oxygens (Figure 1E). These results strongly suggest HglS proceeds via a mechanism similar to HMS and is therefore the third identified member of the two-substrate  $\alpha$ -ketoacid dependent oxygenase (2S- $\alpha$ KAO) family.<sup>17</sup>

Further analysis of the co-crystal structure revealed several other substrate binding residues. Specifically, arginine 74 forms a salt bridge with the distal carboxylate of the 2OA substrate. Additionally, valine 402 and serine 403 form hydrogen bonds with the carboxylate keto group and distal carboxylate group, respectively (Figure 1F). The loop containing residues 402 and 403 shifts approximately 10 Å from the *holo* state to the enzyme-bound state, bringing it into proximity with the substrate. We investigated the importance of Arg74 in determining enzyme specificity by assaying enzyme activity against a panel of  $\alpha$ -ketoacids (Figure 1G). HglS exhibited activity only with 2OA, and displayed no detectable and statistically significant activity above background (a boiled enzyme control) on any other substrate. We therefore concluded that Arg74 participates in favorable substrate binding interactions, but it also appears to influence the strict enzyme substrate specificity. Furthermore, an R74A mutation abolished enzymatic activity (Figure 1H). We additionally probed the importance of the loop bearing Val402 and Ser403 on substrate binding. Mutating Val402 to a proline residue to disrupt the hydrogen bond observed between 2OA and the amide backbone of Val402 significantly decreased enzyme activity, suggesting that while this residue is not essential for turnover, it likely contributes to substrate binding affinity (Figure 1H). However, it is also possible that the

observed decrease in activity is an artefact of disrupted protein folding due to the introduction of a proline residue.



**Figure 1 - Structural and Biochemical Analyses of *P. putida* HgIS.** **A)** Ribbon diagram of the HgIS crystal structure. The left image shows the active site entrance, containing the metal cofactor (red sphere) and metal-coordinating residues (orange) located in the central  $\beta$ -sheet protein domain. A 180° rotation on the right shows the overall protein structure. **B)** Overlay of *holo* (cyan) and substrate-bound (green) structures of HgIS displaying the enzyme active site and the ionic interaction between Arg74 and the distal carboxylate group of the 2OA substrate. The nickel bound within each structure is displayed as a red (*holo* structure) or purple (substrate-bound structure) sphere. **C)** Proposed reaction mechanism of HgIS. O<sub>2</sub>-derived oxygens are shown in red. **D)** Oxygen to 2OA stoichiometry of the HgIS reaction. Dissolved oxygen concentration (y-axis) was measured with a Clark-type oxygen probe. Either 200  $\mu$ M (red line) or 100  $\mu$ M (blue line) of 2OA was added to initiate reaction, resulting in equimolar oxygen consumption. Dotted lines represent the expected final O<sub>2</sub> concentration when 100  $\mu$ M (blue line) or 200  $\mu$ M (red line) 2OA is consumed. **E)** LC-HRMS extracted ion chromatograms (EICs) showing labeled <sup>18</sup>O incorporation into the 2-hydroxyglutarate product of HgIS. On the left (black lines) are EICs for ions with m/z 147.029, representing 2-hydroxyglutarate containing <sup>16</sup>O. On the right (red lines) are EICs for ions with m/z 151.022, representing 2-hydroxyglutarate containing two <sup>18</sup>O atoms. A control reaction performed under ambient O<sub>2</sub> conditions (n=3) is compared to a reaction performed under a <sup>18</sup>O<sub>2</sub> atmosphere (n=3) and a 2-hydroxyglutarate standard. **F)** Overlay of *holo* (green) and substrate-bound (cyan) structures of HgIS displaying the enzyme active site and the interaction of Val402 and Ser403 with the 2-oxoadipate substrate. The loop containing Val402 and Ser403 is shown in the *holo* (dark green) and substrate-bound (dark blue) states. **G)** Reaction rates of HgIS with different 2-oxoacid substrates: 2-oxoglutarate (2OG), 2-oxoadipate (2OA), 2-oxopimelate (2OP), 2-oxobutyrate (2OB), 2-oxovalerate (2OV), 2-oxohexanoate (2OH), 2-oxooctanoate (2OO). Error bars represent 95% confidence intervals, n=3. **H)** Reaction rates of WT HgIS (black dashes), R74A (blue dashes) and V402P (orange dashes) mutants measured by an enzyme coupled decarboxylation assay with 2OA as a substrate, n=3.

## Functional residues of HgIS are highly conserved across the DUF1338 family and suggest a common physiological role

Based on the mechanistic information gleaned from our HgIS biochemical and structural characterization, we sought to propose potential functions for other DUF1338 family members. Previously we showed that DUF1338 family proteins are widely distributed across several domains of life, while others have demonstrated that DUF1338 protein coding sequences are present in the majority of sequenced plant genomes.<sup>8,9</sup> Therefore we analyzed all known DUF1338 domain-containing protein amino acid sequences for the key catalytic and substrate-binding residues we identified in HgIS.

First, we queried all DUF1338 proteins within the Pfam database for the catalytic and metal-binding residues identified in HgIS. Of the 2417 unique DUF1338-containing proteins found in the Pfam PF07063 family, 86% possess the conserved “HHE” metal-binding triad (Table S2). This strong conservation is consistent across domains of life, with 85% of DUF1338 proteins in bacteria, 92% in fungi, and 81% in plants possessing the metal-binding residues. Further analysis revealed arginine 74, which dictates HgIS substrate specificity, is also highly conserved across the family. Of the homologs with conserved HHE triads, 100% of fungal proteins, 99.7% of bacterial proteins, and 93.4% of plant proteins maintain an arginine at this position (Table S2). Therefore, though there is little primary sequence identity between bacterial homologs and plant homologs (Figure S5, Figure S6), the critical residues coordinating the Fe(II) cofactor and the carboxylate-coordinating arginine are highly conserved. Given the substrate specificity exhibited by the *P. putida* DUF1338 protein HgIS, we hypothesized that the uncharacterized homologs maintaining the metal-binding residues and conserved arginine would also use 2OA as a substrate.

To test this hypothesis, we examined the known metabolic pathways with a 2OA intermediate in organisms containing DUF1338 proteins. Plant lysine catabolism involves a 2OA intermediate, though the subsequent steps are undetermined.<sup>7</sup> Notably, nearly all plant and algal genomes within GenBank encode DUF1338 family proteins (Table S2). Previous work to understand plant lysine catabolism suggested the missing steps proceed through the 2-ketoglutarate dehydrogenase complex (2KGD), forming glutaryl-Coenzyme A (Figure S7).<sup>7</sup> Transcript correlation analysis in *Arabidopsis thaliana*, however, revealed that the transcription of known lysine catabolic enzymes was not correlated with 2KGD expression; rather, their transcription is highly correlated with a DUF1338 protein AT1G07040 (Table 1). Additionally, in *Oryza sativa*, mutants of the DUF1338 protein FLO7 display a floury starch phenotype that resembles the maize *opaque2* phenotype which is known to cause lysine accumulation in mutant kernels.<sup>3,5,6,9</sup> These transcriptomic and phenotypic correlations suggest that, like HgLS in *Pseudomonas putida*, DUF1338 proteins are involved in the plant lysine degradation pathway.

During seed maturation, lysine catabolism is required for normal amyloplast development<sup>19</sup>. In rice, *FLO7* localizes to the amyloplast via a N-terminal chloroplast localization signal peptide.<sup>9</sup> While the *A. thaliana* homolog AT1G07040 also has a predicted chloroplast localization sequence, only 117 of 202 plant proteins had predicted chloroplast or mitochondria localization tags (Table S2). Many predicted non-localized proteins are isoforms of loci that also have putatively localized isoforms, though phylogenetic analysis also revealed a distinct non-localized DUF1338 protein clade predominantly within *Brassicaceae* (Figure S8). Notably, this non-localized subclade was enriched for proteins maintaining the HHE coordination triad but that possessed either a glutamine or methionine at the position

corresponding to Arg74 of HgLS (Figure S8). As most plant homologs contain a localization tag, we predicted these homologs would likely function similar to rice *FLO7*.

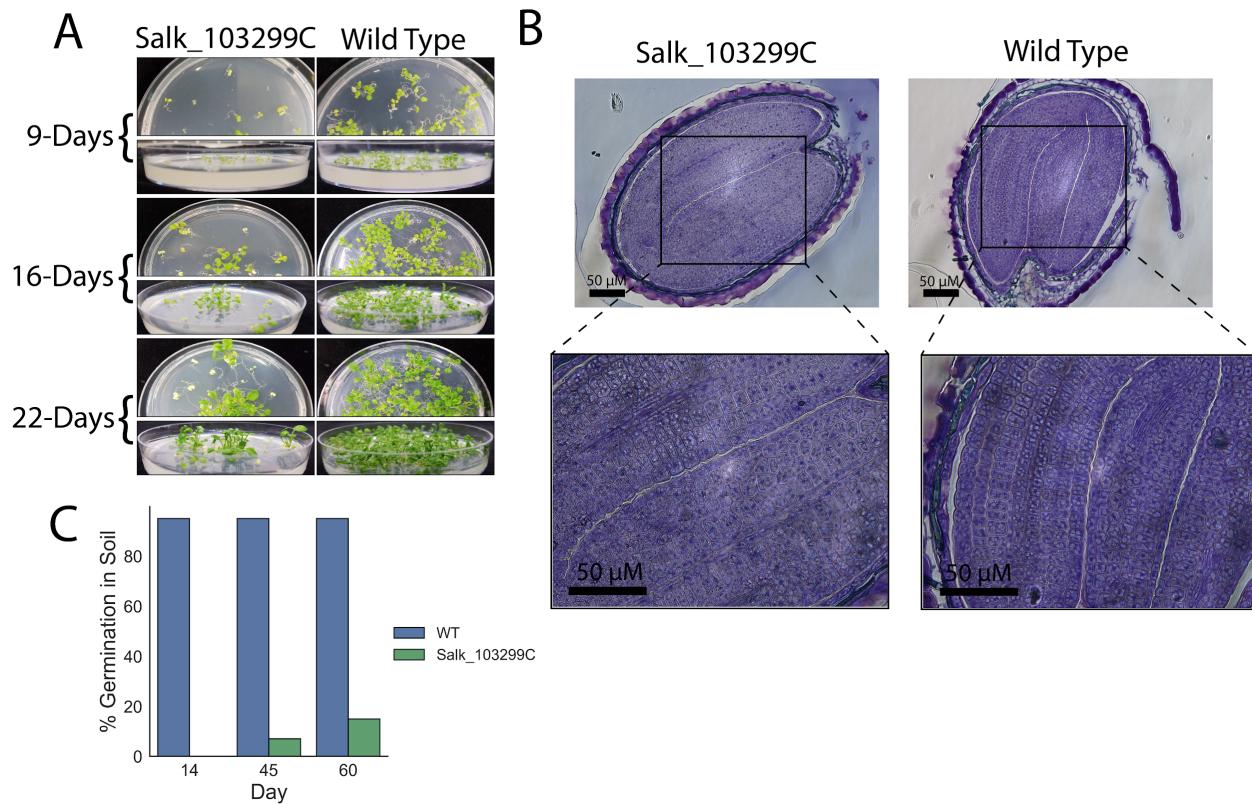
We therefore tested whether mutants of the DUF1338 protein AT1G07040 from the model plant *A. thaliana* exhibited a phenotype resembling other plant lysine-accumulating mutants. When grown on germination media, Salk\_103299C mutants (which have confirmed homozygous T-DNA insertions in the AT1G07040 locus) displayed significantly delayed germination compared to wild type seedlings (Figure 2A). AT1G07040 mutants that germinated also displayed compromised development and appeared chlorotic with impaired growth (Figure 2A). This phenotype was recapitulated when seeds were germinated in soil. After a 14 day incubation, no Salk\_103299C seeds germinated whereas 95% of wild-type seeds had germinated (n=96) (Figure 2C). After 45 days, only 7% of mutant seeds had germinated and after 60 days, only 15%. This phenotype is consistent with previous work showing lysine accumulation in *A. thaliana* seeds resulted in significantly delayed germination.<sup>19,20</sup> Toluidine blue staining of sectioned mutant seeds also suggested that oil body formation was altered (Figure 2B). The altered morphology of the oil body, serving as the primary carbon storage unit in *A. thaliana*, is analogous to the altered starch granule formation observed in rice *FLO7* mutants.<sup>9</sup> The aberrant phenotype in both *A. thaliana* and rice mutants suggest that DUF1338 homologs play a critical role in embryo development both in monocots and dicots.

Reaction	Enzyme	AT Acc. No.	AT1G070 40	AT5G657 50
1	Lysine-2KG reductase/saccharopine dehydrogenase	AT4G3315 0	3	1432.7
2	Aldehyde Dehydrogenase	AT1G5410	1.8	954.6s

		0		
3	Gamma-aminobutyrate transaminase	AT3G2220 0	6.9	1955.9
4	2-Oxoglutarate dehydrogenase complex	AT5G6575 0*	785.2	NA
5	Isovaleryl-coA dehydrogenase	AT3G4530 0	2.4	>2000
6	D-2-Hydroxyglutarate dehydrogenase	AT4G3640 0	15.5	>2000
7	D-2-Hydroxyglutarate synthase	AT1G0704 0	NA	785.2

**Table 1 - ATTED-II correlation of known lysine catabolic genes in *A. thaliana* with HgIS homolog**

**AT1G07040 and the 2KGD complex.** Enzymes catalyzing reactions from Figure S7 were analyzed for their coexpression with either HgIS or 2KGD in *A. thaliana* and reported by their Mutual Rank Index.<sup>21</sup> HgIS was highly correlated with all known lysine catabolic enzymes (column AT1G07040), while there was little correlation with the 2KGD complex (column AT5G65750). AT5G65750\* represents the 2KGD complex E1 subunit only.



**Figure 2 - Germination rates of WT *A. thaliana* and Salk\_103299C mutants. A)** Arabidopsis seedlings germination and growth on germination medium at 9, 16, or 22 days after incubation. **B)** Toluidine blue stained histopathology of mature seeds from Salk\_103299C mutant (left) and wild type seeds of *A. thaliana* (right). Sections are shown at 20x (top panels) and 40x (bottom panels) magnification. **C)** Germination rates of Salk\_103299C relative to WT *A. thaliana*. Salk\_103299C mutants showed delayed germination relative to WT *A. thaliana*. No mutants germinated 14 days after planting while 95% of WT seeds had germinated, n=96. 60 days after planting only 15% of mutant seeds had germinated.

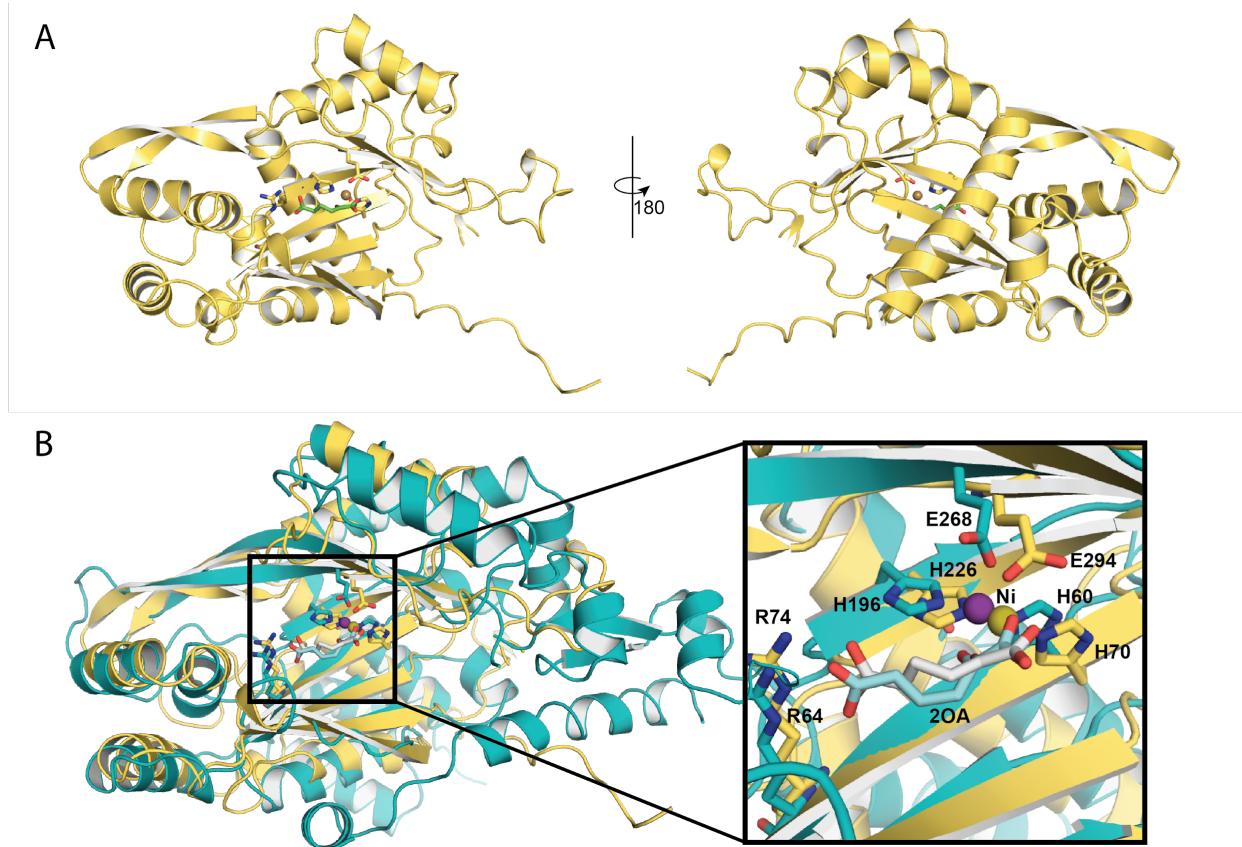
### The biochemical activity and structure of DUF1338 proteins is conserved in plant and bacterial homologs

Transcriptional correlation analysis and the observed phenotypes in *A. thaliana* Salk\_103299C mutants strongly suggested a role for plant DUF1338 homologs in lysine

catabolism. We hypothesized that these proteins perform the same reaction as HglS, converting 2OA to 2HG. To test this, we biochemically characterized the DUF1338 homologs from *A. thaliana* and *O. sativa* as well as an additional distantly related *Escherichia coli* homolog. Soluble variants of plant proteins were constructed by removing the predicted N-terminal localization peptide. As expected, the *E. coli* homolog YdcJ and the plant homologs AT1G07040 and FLO7 catalyzed the conversion of 2OA to 2HG which was confirmed by *in vitro* assays analyzed using high resolution LC-MS (Figure S9A). Additionally, the FLO7 homolog displayed kinetic parameters similar to HglS, with a  $K_m$  of 0.55 mM and a  $V_{max}$  of 0.89 mM/min/ $\mu$ M enzyme (Figure S9B).

Additionally, the structural conservation between bacterial and eukaryotic DUF1338 proteins was compared by obtaining a crystal structure of FLO7. Initial crystallization screens resulted in crystals which were soaked with 2OA, producing diffraction data used to solve the substrate-bound crystal structure at 1.85 Å resolution. The FLO7 crystal structure, like the other DUF1338 proteins, displayed the conserved central VOC fold containing the active site and metal-binding center (Figure 3A). Comparison of the HglS and FLO7 structures revealed that orientation of the metal-coordinating residues and 2OA in both structures were nearly identical (Figure 3B) even though the proteins display low (~15%) sequence identity (Figure S5). In addition, the FLO7 structure also contains a substrate-binding arginine, Arg64, located in nearly the same position as Arg74 of HglS. However, unlike Val402 and Ser403 of HglS, FLO7 does not possess any other residues interacting with the substrate. The results of our biochemical and structural studies of FLO7 manifest the remarkable conservation of the fold, active site architecture, and mechanism among DUF1338 proteins across different domains of life. Consequently, we predict that all DUF1338 proteins containing the conserved 2OA-interacting

arginine likely catalyze the decarboxylation and hydroxylation of 2OA to form 2HG. These hydroxyglutarate synthases form a new two-substrate  $\alpha$ -keto acid dependent oxygenase (2S- $\alpha$ KAO) subfamily lacking significant sequence identity to hydroxymandelate synthase or 4-hydroxyphenyl pyruvate dioxygenase and acting on a distinct substrate.



**Figure 3. A)** FLO7 crystal structure displayed as cartoon sheets and helices. Metal-coordinating and substrate-binding residues are shown as sticks, while the 2-oxoadipate substrate is shown as green sticks. The nickel bound within the active site is shown as a tan sphere. **B)** Overlay of HgIS (teal) and FLO7 (yellow) structures showing metal-coordinating residues, substrate-binding residues, the active site metal, and the 2-oxoadipate substrate. Right inset: HgIS (teal) and FLO7 (yellow) active sites displaying conserved residues and substrate-binding mode. The 2OA substrates are colored light blue (HgIS) and white (FLO7), and the active site bound nickel shown as spheres and colored blue (HgIS) and olive (FLO7).

## DISCUSSION

Here, we present the structural and mechanistic analysis of HgIS, the first DUF1338 protein family member to be assigned a function. In a previous report, we showed that HgIS converts 2OA to 2HG. The HgIS chemical mechanism remained ambiguous however, as this conversion involves two apparent enzymatic steps: decarboxylation and hydroxylation. Here we report the high resolution crystal structures of *holo*- and substrate-bound enzyme aiding the proposal of an enzyme mechanism. In addition, we present the substrate-bound structure of the plant homolog, FLO7.

Several HgIS unique features were used to propose an enzymatic reaction mechanism. A VAST structural similarity search revealed the central VOC superfamily  $\beta$ -sheet fold, and metal-binding residues of HgIS were positioned similarly to the corresponding HMS and HPPD residues (Table S1). Both HMS and HPPD catalyze an intramolecular decarboxylation-hydroxylation reaction but share little sequence identity with HgIS.<sup>13</sup> Furthermore, a 2OA-bound HgIS co-crystal structure revealed bidentate coordination of the 2OA  $\alpha$ -keto group by the metal center as a key step in the enzymatic mechanism. These features were remarkably conserved in the FLO7 co-crystal structure, while the  $\alpha$ -helices surrounding central domain diverge.

The VOC fold is conserved across the protein superfamily, yet the specific orientation of  $\beta$ -sheets in HgIS, FLO7, HMS, and HPPD appears to be most conserved in the intramolecular 2S- $\alpha$ KAO subfamily. With only seven  $\beta$ -sheets composing the central barrel-like domain, the HgIS structure diverges slightly from the canonical VOC fold which contains eight  $\beta$ -sheets.

The central VOC domains of HgIS, FLO7, HMS, and HPPD are distinct from that of the mechanistically similar  $\alpha$ -ketoglutarate ( $\alpha$ KG) dependent dioxygenases which possess a double-

stranded  $\beta$ -helix (jelly roll) fold containing the metal-binding center and active site.<sup>22</sup> However, the orientation of metal-binding residues and bound substrate are understandably homologous given the similar chemical mechanisms of the enzyme families which we verified through several biochemical experiments. Whether the shared general mechanism of the 2S- $\alpha$ KAOs and  $\alpha$ KG dependent dioxygenases is a product of convergent evolution remains unclear.

While HMS and HPPD show significant sequence identity and share the same substrate, the DUF1338 family proteins are not homologous to HMS or HPPD, nor do they share significant sequence identity with each other (~15% identity between Hg1S and FLO7). Low sequence conservation is a general VOC superfamily feature, but the low sequence identity between members catalyzing the same reaction is less common.<sup>17</sup> Despite low primary sequence conservation, the biochemical assays of mutants, structural analysis, and bioinformatics reported herein show that very few residues are essential for enzyme turnover in the hydroxyglutarate synthase family. It is possible that the  $\alpha$ -helices surrounding the central VOC domain have been selected for other attributes, such as mediating protein-protein interactions as observed in the  $\alpha$ KG-dependent dioxygenases.<sup>22</sup>

Previous work has shown DUF1338 family proteins are widespread in bacteria and fungi, but they are especially prevalent in plants.<sup>8,9</sup> Our work here shows that the arginine residue that mediates the 2OA specificity of Hg1S is highly conserved across these homologs. It is therefore likely that all DUF1338 proteins with the conserved arginine catalyze the decarboxylation and hydroxylation of 2OA to form 2HG. The enzyme's substrate, 2OA, is primarily known as a lysine catabolism intermediate, suggesting that DUF1338 enzymes participate in similar biochemical processes across domains of life

(<http://modelseed.org/biochem/compounds/cpd00269>). Previous plant phenotypic studies support

this hypothesis; a rice study showed DUF1338-containing protein *FLO7* disruption produced a crystalline starch phenotype within the amyloplast,<sup>9</sup> similar to the lysine-accumulating *opaque2* maize mutants.<sup>23</sup>

DUF1338 protein AT1G07040 expression in *A. thaliana* is highly correlated with other known lysine catabolism enzymes. In this work, we further support this claim by showing that when AT1G07040 is disrupted, *A. thaliana* seedlings have significantly compromised germination ability. Previous work showed *A. thaliana* lysine-accumulating mutants have delayed germination rates due to unfavorable TCA cycle effects, suggesting that AT1G07040 could be involved in lysine degradation.<sup>19,20</sup> Furthermore, by assaying the *A. thaliana* and rice enzymes *in vitro*, we show that AT1G07040 and FLO7 catalyze the transformation of 2OA to 2HG. While both compounds were previously identified as plant lysine catabolism intermediates, no clear link between the two molecules had been demonstrated.<sup>24</sup> Finally, histopathological examination of mutant AT1G07040 seeds showed aberrant oil body formation. While previous experiments showed that the disruption of lysine utilization compromises the ability of seeds to store carbon in monocots such as rice and maize, our results show this phenomenon also occurs in dicots such as *Arabidopsis*. Given their near ubiquitous conservation and highly specific biochemical function we find it likely that DUF1338 proteins localized to chloroplasts catalyze the last missing step in lysine catabolism of all green plants.

While the majority of plant Hg1S homologs have chloroplast localization tags, some lack any predicted signal peptide. This is especially prevalent in the *Brassicaceae*, where many species appear to have non-localized Hg1S paralogs. The majority of these paralogs retain arginine as their specificity residue, but they may have roles in pathways beyond lysine

catabolism as the *A. thaliana* paralog AT1G27030 (lacking a localization sequence) shows no expression correlation with known lysine catabolic genes (data not shown).

Recent studies show that the lysine-derived intermediates pipecolate and N-hydroxy-pipecolate can initiate systemic acquired resistance (SAR) in plants.<sup>25,26</sup> SAR, a global response, grants lasting broad-spectrum disease protection in uninfected tissue.<sup>27</sup> In bacteria pipecolate is often catabolized to 2OA, suggesting that HglS homologs should receive future attention when studying pipecolate metabolism in plants. In addition to the non-localized paralogs retaining the conserved arginine, multiple *Brassicaceae* paralogs have altered residues at this position, harboring either methionine or asparagine. These paralogs likely catalyze the decarboxylation and hydroxylation of substrates other than 2OA. Further research discerning the physiological and biochemical functionality of “mutant” paralogs is warranted.

DUF1338 proteins are also widely distributed in both *Ascomycota* and *Basidiomycota* fungi, though the model fungi *Saccharomyces cerevisiae* or *Neurospora crassa* lack homologs. Within the fungal homologs examined here, all proteins containing a HHE metal binding triad also maintained the arginine specificity residue suggesting that all function on a 2OA substrate. Unfortunately, almost no fungal catabolic lysine pathways are fully characterized genetically or biochemically.<sup>28</sup> Multiple studies, however, suggested 2OA is a likely lysine catabolism intermediate in *Pyriculuria oryzae* and *Candida albicans*.<sup>29,30</sup> Moreover, while *P. oryzae* possesses a DUF1338 homolog, *C. albicans* does not. This implies *C. albicans* and other fungi may utilize a catabolic route similar to mammals in which 2OA is converted to glutaryl-CoA via 2-oxoadipate dehydrogenase.<sup>31</sup> Future studies are required to examine whether fungal HglS homologs also play a role in fungal lysine catabolism.

Of the over 2000 bacterial Hgls homologs examined here that retain the HHE triad, over 99% maintained arginine as the specificity-conferring residue, suggesting widespread DUF1338 enzymatic activity conservation in prokaryotes. Previous work in *P. putida* and published fitness data in *Pseudomonas fluorescens* and *Sinorhizobium meliloti* provide evidence of a conserved physiological function as well.<sup>8,32,33</sup> *E. coli* was recently shown to possess a non-ketogenic lysine catabolic route via a glutarate hydroxylase, supplementing degradation routes to cadaverine.<sup>34,35</sup> *E. coli* also possesses a Hgls homolog, YdcJ, which we showed has identical activity to Hgls. Additionally, the *E. coli* enzyme structure was solved (PDB ID 2RJB) and shows the same conserved VOC fold, metal-binding motif, and arginine-binding residues as the Hgls and FLO7 structures reported herein. Future work will elucidate whether, like *P. putida*, *E. coli* possesses multiple lysine catabolism routes or whether 2OA functions in other physiological processes.

Low cereal and legume lysine content produces protein-energy malnutrition in 30% of the developing world population.<sup>3,36-39</sup> Increasing lysine content in staple crops will require both lysine overproduction and catabolic pathway elimination.<sup>38</sup> However, lysine metabolism changes in maize, rice, and soybeans resulted in low germination rates, abnormal endosperm, and reduced grain weights.<sup>3,23,40-42</sup> We hope that the more complete understanding of lysine catabolism elucidated here will help resolve causes of pleiotropic effect in plants and aid in the development of stable high-lysine crops to combat malnutrition globally.

## Methods

### Media, Chemicals, and Strains

Routine bacterial cultures were grown in Luria-Bertani (LB) Miller medium (BD Biosciences, USA). *E. coli* was grown at 37 °C. Cultures were supplemented with carbenicillin (100 mg/L, Sigma Aldrich, USA). All compounds with the exception of 2-oxohexanoic acid

were purchased through Sigma Aldrich. All bacterial strains and plasmids used in this work are listed in Table S4 and are available through the public instance of the JBEI registry. (<https://public-registry.jbei.org/>). *Arabidopsis thaliana* mutants were obtained from the Salk collection.

### DNA Manipulation

All plasmids were designed using Device Editor and Vector Editor software, while all primers used for the construction of plasmids were designed using j5 software.<sup>43-45</sup> Plasmids were assembled via Gibson Assembly using standard protocols,<sup>46</sup> or Golden Gate Assembly using standard protocols.<sup>47</sup> Plasmids were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Site directed mutants were created as previously described.<sup>48</sup> The geneblock for the *E. coli* codon optimized *O. sativa* was purchased through Integrated DNA Technologies (IDT, Coralville, IA). Arabidopsis cDNA was used to amplify AT1G07040.

### Protein purification

A 5 mL overnight culture of *E. coli* BL21 (DE3) containing the expression plasmid was used to inoculate a 500 mL culture of LB. Cells were grown at 37 °C to an OD of 0.6 then induced with Isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. The temperature was lowered to 30 °C and cells were allowed to express for 18 hours before being harvested via centrifugation. Cell pellets were stored at -80 °C until purification. For purification, cell pellets were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, 8% glycerol, pH 7.5) and sonicated to lyse cells. Insolubles were pelleted via centrifugation (30 minutes at 40,000xg). The supernatant was applied to a fritted column containing Ni-NTA resin (Qiagen, USA) which had been pre-equilibrated with several

column volumes of lysis buffer. The resin was washed with lysis buffer containing 50 mM imidazole, then the protein was eluted using a stepwise gradient of lysis buffer containing increasing imidazole concentrations (100 mM, 200 mM, and 400 mM). Fractions were collected and analyzed via SDS-PAGE. Purified protein was dialyzed overnight at 4 °C against 50 mM HEPES pH 7.5, 5% glycerol.

### Crystallization

An initial crystallization screen was set up using a Phoenix robot (Art Robbins Instruments, Sunnyvale, CA) using the sparse matrix screening method.<sup>49</sup> Purified Hg1S was concentrated to 20 mg/mL and Flo7 was concentrated to 10 mg/mL prior to crystallization using the sitting drop method in 0.4 µL drops containing a 1:1 ratio of protein sample to crystallization solution. For Hg1S, the crystallization solution consisted of 0.2 M Ammonium Fluoride and 20 % PEG 3,350, while the crystallization solution for Flo7 contained 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 5.6 and 1.8 M Lithium sulfate monohydrate. Crystals were transferred to crystallization solution containing 20 % glycerol prior to flash freezing in liquid nitrogen.

### X-ray data collection and model refinement

X-ray diffraction data for Hg1S was collected at the Berkeley Center for Structural Biology on beamline 5.0.2 and 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Lab. Diffraction data for Flo7 was collected at the Stanford Synchrotron Radiation Lightsource on beamline 12-2. The Hg1S and Flo7 structures were determined by the molecular-replacement method with the program *PHASER*<sup>50</sup> using uncharacterized protein YdcJ (SF1787) from *Shigella flexneri* (PDB ID: 2RJB) and the putative hydrolase (YP\_751971.1) from *Shewanella frigidimarina* (PDB ID: 3LHO) as the search models, respectively. Structure

refinement was performed by *phenix.refine* program.<sup>51</sup> Manual rebuilding using COOT<sup>52</sup> and the addition of water molecules allowed for construction of the final model. The R-work and R-free values for the final models of all structures are listed in Table SX. Root-mean-square deviations from ideal geometries for bond lengths, angles, and dihedrals were calculated with Phenix.<sup>53</sup> The overall stereochemical quality of the final models was assessed using the MolProbity program.<sup>54</sup> Structural analyses were performed in Coot,<sup>52</sup> PyMOL,<sup>55</sup> and UCSF Chimera.<sup>56</sup>

#### Enzyme Kinetics and O<sub>2</sub> Consumption

Enzyme coupled decarboxylation assays were carried out as previously described.<sup>57</sup> Reaction mixtures contained 100 mM Tris-HCl (pH 7), 10 mM MgCl<sub>2</sub>, 0.4 mM NADH, 4 mM phosphoenol pyruvate (PEP), 100 U/mL pig heart malate dehydrogenase (Roche), 2U/mL microbial PEP carboxylase (Sigma), and 10 mM 2OA. Reactions were initiated by the addition of purified HgIS or boiled enzyme controls, and absorbance at 340 nm was measured via a SpectraMax M4 plate reader (Molecular Devices, USA).

Initial rate measurements were directly recorded monitoring the consumption of O<sub>2</sub> using a Clarke-type electrode (Hansatech Oxygraph). Reaction mixtures containing FeSO<sub>4</sub> (10 μM) and 2-oxoadipic acid (10-200 μM) in 100 mM Tris pH 7.0 were allowed to equilibrate to room temperature determined by a stable O<sub>2</sub> concentration reading of 240 μM. Addition of purified *apo* enzyme (100 nM) to the sealed reaction vial initiated the reaction. Initial rates were determined from the linear portion of consumption of O<sub>2</sub> corresponding to up to 10% consumption of the limiting reactant. No burst or lag phases were observed. All assays were performed in triplicate.

O<sub>2</sub> consumption measurements used to determine the reaction stoichiometry were also measured using a Clarke-type electrode. Reactions mixtures containing FeSO<sub>4</sub> (10 μM) and 2-oxoadipic acid (100 or 200 μM) in 100 mM Tris pH 7.0 were equilibrate to room temperature ([O<sub>2</sub>]

= 240  $\mu$ M) and monitored for at least 2 minutes prior to the addition of *apo* enzyme (1  $\mu$ M).

Oxygen consumption was monitored until the signal plateaued and the O<sub>2</sub> concentration was stable.

The observed rate was determined by fitting the data to a single exponential decay model. The reaction stoichiometry was determined by taking the ratio of the moles of O<sub>2</sub> consumed and the concentration of the 2-oxoadipic acid present in the reaction. Each reaction condition was performed in triplicate.

#### Oxygen Labelling Experiments

All reagents were exhaustively purged with argon on a Schlenk link to remove <sup>16</sup>O<sub>2</sub>. Anaerobic buffer was subsequently saturated with <sup>18</sup>O<sub>2</sub> via gentle bubbling with <sup>18</sup>O<sub>2</sub> (Sigma, 99 atom % <sup>18</sup>O). Reaction mixtures containing enzyme (1  $\mu$ M), 2-oxoadipic acid (1 mM), FeSO<sub>4</sub> (10  $\mu$ M) and <sup>18</sup>O<sub>2</sub> saturated buffer (1.2 mM) were mixed in anaerobic sealed reaction vials using gastight syringes. The reaction was initiated by the addition of 2-oxoadipic acid. The headspace was filled with <sup>18</sup>O<sub>2</sub> gas. Reactions were incubated at room temperature for 2 hours before being quenched with an equal volume of methanol. Control experiments, replacing <sup>18</sup>O<sub>2</sub> with <sup>16</sup>O<sub>2</sub>, were run in parallel. Quenched reaction mixtures were analyzed by LC-MS.

#### LCMS Analysis

All *in vitro* reactions to be analyzed via LC-MS were quenched with an equal volume of ice cold methanol and stored at -80C until analyses. Detection of 2OA and 2HG were described previously.<sup>8</sup> Briefly, HILIC-HRMS analysis was performed using an Agilent Technologies 6510 Accurate-Mass Q-TOF LC-MS instrument using positive mode and an Atlantis HILIC Silica 5  $\mu$ m column (150 x 4.6 mm) with a linear of 95 to 50% acetonitrile (v/v) over 8 minutes in water with 40 mM ammonium formate, pH 4.5, at a flow rate of 1 mL minute<sup>-1</sup>.

### Synthesis of 2-oxohexanoic acid

Synthesis of 2-oxohexanoic acid was carried out as described previously.<sup>58</sup> Briefly, a solution of the Grignard reagent, prepared from 1- bromobutane (500 mg, 3.68 mmol) and the suspension of magnesium (178 mg, 7.33 mmol) in THF (5 mL) was added dropwise under N<sub>2</sub> atmosphere to a solution of diethyloxalate (487 mg, 3.33 mmol) in THF (4 mL) at -78 °C. After the addition was complete, the reaction mixture was stirred at -78 °C for an additional 5 h. The reaction was quenched with 2 N HCl, the aqueous layer was extracted with ether, and the combined organic layer was washed with brine, dried over MgSO<sub>4</sub>, and evaporated. The crude product was dissolved in acetic acid (20 mL) and conc. HCl (5 mL). After 11 h, the reaction was concentrated directly, and the residue was purified by distillation under reduced pressure to give the pure product (203 mg, 43%) as a colorless oil. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) 2.92 (t, 2H), 1.81 – 1.54 (m, 2H), 1.48 – 1.28 (m, 2H), 0.92 (t, 3H) (Figure S10).

### Plant Growth

*Arabidopsis thaliana* DUF1338 mutant, Salk\_103299C, seeds were ordered from the Arabidopsis Biological Resource Center (Columbus, OH, USA). Seeds were surface sterilized with 70% EtOH for three minutes, followed by a ten-minute submersion in 10% bleach solution, then rinsed with sterile water three times. For soil germination, seeds were planted in Premier Pro-Mix mycorise pro soil, two seeds per well in a twenty-four well tray. Trays with seeds were then stratified at 4°C for 3 days, then grown in a Percival-Scientific growth chambers at 22°C in 10/14-hour light/dark short-day cycles with 60% humidity. After the formation of the full rosette, plants were genotyped to confirm homozygosity for the mutation of interest. Primers were designed using the automated method from <http://signal.salk.edu/cgi-bin/tdnaexpress>, and

genotyping was done in accordance with a previously described protocol (Figure S10).<sup>59</sup> After confirmation of homozygous mutants, plants were moved into individual wells and transferred to long-day conditions, 22°C in 16/8-hour light/dark cycles with 60% humidity, to induce flowering for seed collection. For germination on synthetic media, sterilized seeds were plated on solid media supplemented with ½ Murashige and Skoog media base, 5% sucrose, 0.8% agar, and 10uM gibberellic acid to induce germination. Plates with seeds were stratified at 4°C for 3 days then transferred to a Percival-Scientific growth chambers at 22°C in 10/14-hour light/dark short-day cycles with 60% humidity.

### Histopathology

Seeds were imbibed in distilled water for 2h with gently shaking and a small hole was cut in the seed coat with a surgical scalpel to aid resin infiltration. Seeds were fixed in 4% formaldehyde (Electron Microscopy Sciences) in 50mM PIPES buffer (pH 7) with gently pulling under vacuum for 10 min. Seeds were left in fixative overnight at 4°C, dehydrated in an ethanol gradient, and infiltrated with Technovit 7100 plastic resin (Electron Microscopy Sciences) according to manufacturer's instructions. Resin was infiltrated under gentle vacuum for 20 min followed by rotation at 4°C overnight. Seeds were embedded in beam capsules and 4 µm thick sections were cut on an MR2 manual rotary microtome (RMC Boeckeler) using a glass knife. Sections were stained with 0.02% (w/v) Toluidine blue-O for 30 sec, then rinsed and mounted in distilled water. Images were captured on a Leica DM6B microscope (Leica Biosystems Inc. Buffalo Grove, IL) equipped with a Leica DMC 4500 color camera.

### Bioinfomatics

For mining of DUF1338 homologs in plants all the proteins from completed genomes with protein predictions available by september 2019 were retrieved from the GenBank FTP site.

The homologs were searched in each proteome with BlastP using Flo7 as query with a bit score cut-off of 150 and an e-value cut-off of E-12. The retrieved sequences were aligned using muscle<sup>60</sup>, and trimmed using jalview<sup>61</sup>. The multiple sequence alignment was used for phylogenetic reconstruction using IQ tree, the best amino acid substitution model was selected with ModelFinder implemented in IQtree<sup>62</sup>, branch support was calculated using 10000 bootstrap generations

Sequences of DUF1338 homologs were downloaded from Pfam (<https://pfam.xfam.org/family/PF07063>). To compare the 3D structure of Hg1S with other protein structures we used Vector Alignment Search Tool (VAST).<sup>13</sup> All alignments were done using the MAFFT-LINSI algorithm,<sup>63</sup> and alignments were compared to secondary structures and visualized using Easy Sequencing in PostScrip (<http://escript.ibcp.fr>).<sup>64</sup> Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.<sup>56</sup> Custom Python scripts were developed to calculate conservation of “HHE” metal binding triads and R74 residues. Calculation of Shannon-Entropy for DUF1338 sequence conservation was carried out using the python library Protein Dynamics and Sequence Analysis (ProDy).<sup>65,66</sup> Co-expression analysis of *A. thaliana* transcripts was performed using the ATTED-II database.<sup>21</sup>

## Acknowledgements

We would like to thank Johan Jaenisch for generously providing *A. thaliana* cDNA. Python code to analyze kinetics data was provided by Sam Curran. This work was part of the DOE Joint BioEnergy Institute (<https://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and was part of the

Agile BioFoundry (<http://agilebiofoundry.org>) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical Institute. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

## Contributions

Conceptualization, M.G.T., J.M.B, and J.H.P.; Methodology, M.G.T., J.M.B, J.H.P., J.F.B., P.C.M., E.E.K.B, J.H, W.M., M.B.,Y.L.; Investigation, M.G.T., J.M.B, J.H.P., V.T., J.H, M.B., W.M., L.J.W., T.B.,W.S., R.W.H, C.B.E, E.E.K.B.,Y.L.; Writing – Original Draft, M.G.T, J.M.B. J.H.P.; Writing – Review and Editing, All authors.; Resources and supervision H.V.S., P.M.S., M.M., P.D.A., and J.D.K.

## Bibliography

1. Galili, G. New insights into the regulation and functional significance of lysine metabolism in plants. *Annu. Rev. Plant Biol.* **53**, 27–43 (2002).
2. Galili, G. & Höfgen, R. Metabolic engineering of amino acids and storage proteins in plants. *Metab. Eng.* **4**, 3–11 (2002).
3. Galili, G. & Amir, R. Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnol. J.* **11**, 211–222 (2013).
4. Yang, Q.-Q. *et al.* Biofortification of rice with the essential amino acid lysine: molecular characterization, nutritional evaluation, and field performance. *J. Exp. Bot.* **67**, 4285–4296 (2016).
5. Hartings, H., Lauria, M., Lazzaroni, N., Pirona, R. & Motto, M. The Zea mays mutants opaque-2 and opaque-7 disclose extensive changes in endosperm metabolism as revealed by protein, amino acid, and transcriptome-wide analyses. *BMC Genomics* **12**, 41 (2011).
6. Locatelli, S., Piatti, P., Motto, M. & Rossi, V. Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm development. *Plant Cell* **21**, 1410–1427 (2009).
7. Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L. & Braun, H.-P. Amino acid catabolism in plants. *Mol. Plant* **8**, 1563–1579 (2015).

8. Thompson, M. G. *et al.* Massively Parallel Fitness Profiling Reveals Multiple Novel Enzymes in *Pseudomonas putida* Lysine Metabolism. *MBio* **10**, (2019).
9. Zhang, L. *et al.* FLOURY ENDOSPERM7 encodes a regulator of starch synthesis and amyloplast development essential for peripheral endosperm development in rice. *J. Exp. Bot.* **67**, 633–647 (2016).
10. Joint Center for Structural Genomics (JCSG). Crystal structure of putative hydrolase (YP\_751971.1) from *Shewanella frigidimarina* NCIMB 400 at 1.80 Å resolution. *TO BE PUBLISHED*
11. Joint Center for Structural Genomics (JCSG). Crystal structure of Putative glyoxalase superfamily protein (YP\_299723.1) from *RALSTONIA EUTROPHA* JMP134 at 1.90 Å resolution. *TO BE PUBLISHED*
12. Seetharaman, J. *et al.* Crystal structure of uncharacterized protein YdcJ (SF1787) from *Shigella flexneri* which includes domain DUF1338. *TO BE PUBLISHED*
13. Madej, T. *et al.* MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res.* **42**, D297-303 (2014).
14. Brownlee, J., He, P., Moran, G. R. & Harrison, D. H. T. Two roads diverged: the structure of hydroxymandelate synthase from *Amycolatopsis orientalis* in complex with 4-hydroxymandelate. *Biochemistry* **47**, 2002–2013 (2008).
15. Serre, L. *et al.* Crystal structure of *Pseudomonas fluorescens* 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. *Structure* **7**, 977–988 (1999).
16. Di Giuro, C. M. L. *et al.* Chiral hydroxylation at the mononuclear nonheme Fe(II) center of 4-(S) hydroxymandelate synthase--a structure-activity relationship analysis. *PLoS ONE* **8**,

- e68932 (2013).
17. He, P. & Moran, G. R. Structural and mechanistic comparisons of the metal-binding members of the vicinal oxygen chelate (VOC) superfamily. *J. Inorg. Biochem.* **105**, 1259–1272 (2011).
  18. Armstrong, R. N. Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry* **39**, 13625–13632 (2000).
  19. Angelovici, R., Fait, A., Fernie, A. R. & Galili, G. A seed high-lysine trait is negatively associated with the TCA cycle and slows down *Arabidopsis* seed germination. *New Phytol.* **189**, 148–159 (2011).
  20. Zhu, X. & Galili, G. Increased lysine synthesis coupled with a knockout of its catabolism synergistically boosts lysine content and also transregulates the metabolism of other amino acids in *Arabidopsis* seeds. *Plant Cell* **15**, 845–853 (2003).
  21. Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y. & Kinoshita, K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant Cell Physiol.* **59**, e3 (2018).
  22. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22**, 691–700 (2012).
  23. Mertz, E. T., Bates, L. S. & Nelson, O. E. Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science* **145**, 279–280 (1964).
  24. Engqvist, M. K. M. *et al.* Plant D-2-hydroxyglutarate dehydrogenase participates in the catabolism of lysine especially during senescence. *J. Biol. Chem.* **286**, 11382–11390 (2011).
  25. Chen, Y.-C. *et al.* N-hydroxy-pipecolic acid is a mobile metabolite that induces systemic

- disease resistance in *Arabidopsis*. *Proc Natl Acad Sci USA* **115**, E4920–E4929 (2018).
26. Wang, C. *et al.* Pipecolic acid confers systemic immunity by regulating free radicals. *Sci. Adv.* **4**, eaar4509 (2018).
  27. Návarová, H., Bernsdorff, F., Döring, A.-C. & Zeier, J. Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity. *Plant Cell* **24**, 5123–5141 (2012).
  28. Zabriskie, T. M. & Jackson, M. D. Lysine biosynthesis and metabolism in fungi. *Nat. Prod. Rep.* **17**, 85–97 (2000).
  29. Wade, M., Thomson, D. M. & Miflin, B. J. Saccharopine: an Intermediate of L-Lysine Biosynthesis and Degradation in *Pyricularia oryzae*. *MICROBIOLOGY* **120**, 11–20 (1980).
  30. Hammer, T., Bode, R. & Birnbaum, D. Occurrence of a novel yeast enzyme, L-lysine epsilon-dehydrogenase, which catalyses the first step of lysine catabolism in *Candida albicans*. *J. Gen. Microbiol.* **137**, 711–715 (1991).
  31. Hallen, A., Jamie, J. F. & Cooper, A. J. L. Lysine metabolism in mammalian brain: an update on the importance of recent discoveries. *Amino Acids* **45**, 1249–1272 (2013).
  32. Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306-15 (2015).
  33. Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**, 503–509 (2018).
  34. Soksawatmaekhin, W., Kuraishi, A., Sakata, K., Kashiwagi, K. & Igarashi, K. Excretion and uptake of cadaverine by CadB and its physiological functions in *Escherichia coli*. *Mol. Microbiol.* **51**, 1401–1412 (2004).
  35. Knorr, S. *et al.* Widespread bacterial lysine degradation proceeding via glutarate and L-2-

- hydroxyglutarate. *Nat. Commun.* **9**, 5071 (2018).
36. Grover, Z. & Ee, L. C. Protein energy malnutrition. *Pediatr. Clin. North Am.* **56**, 1055–1068 (2009).
37. Wenefrida, I., Utomo, H. S., Blanche, S. B. & Linscombe, S. D. Enhancing essential amino acids and health benefit components in grain crops for improved nutritional values. *Recent Pat. DNA Gene Seq.* **3**, 219–225 (2009).
38. Le, D. T., Chu, H. D. & Le, N. Q. Improving nutritional quality of plant proteins through genetic engineering. *Curr. Genomics* **17**, 220–229 (2016).
39. WHO. *Protein and amino acid requirements in human nutrition*. (World Health Organization, 2007).
40. Kawakatsu, T., Wang, S., Wakasa, Y. & Takaiwa, F. Increased lysine content in rice grains by over-accumulation of BiP in the endosperm. *Biosci. Biotechnol. Biochem.* **74**, 2529–2531 (2010).
41. Falco, S. C. *et al.* Transgenic canola and soybean seeds with increased lysine. *Biotechnology (NY)* **13**, 577–582 (1995).
42. Betrán, F. J., Bockholt, A., Fojt, F., Rooney, L. & Waniska, R. Registration of Tx802. *Crop Sci.* **43**, 1891-a (2003).
43. Ham, T. S. *et al.* Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* **40**, e141 (2012).
44. Chen, J., Densmore, D., Ham, T. S., Keasling, J. D. & Hillson, N. J. DeviceEditor visual biological CAD canvas. *J. Biol. Eng.* **6**, 1 (2012).
45. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA assembly design automation software. *ACS Synth. Biol.* **1**, 14–21 (2012).

46. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
47. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
48. Yan, P., Gao, X., Shen, W., Zhou, P. & Duan, J. Parallel assembly for multiple site-directed mutagenesis of plasmids. *Anal. Biochem.* **430**, 65–67 (2012).
49. Jancarik, J. & Kim, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411 (1991).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
51. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
53. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
54. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-83 (2007).
55. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. . at <<https://pymol.org/2/>>
56. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Witkowski, A., Joshi, A. K. & Smith, S. Mechanism of the  $\beta$ -Ketoacyl Synthase Reaction

- Catalyzed by the Animal Fatty Acid Synthase<sup>†</sup>. *Biochemistry* **41**, 10877–10887 (2002).
58. Rapf, R. J. *et al.* Photochemical Synthesis of Oligomeric Amphiphiles from Alkyl Oxoacids in Aqueous Environments. *J. Am. Chem. Soc.* **139**, 6946–6959 (2017).
  59. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
  60. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
  61. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
  62. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
  63. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
  64. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-4 (2014).
  65. Bakan, A. *et al.* Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **30**, 2681–2683 (2014).
  66. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
  67. Grant, S. G., Jesse, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl*

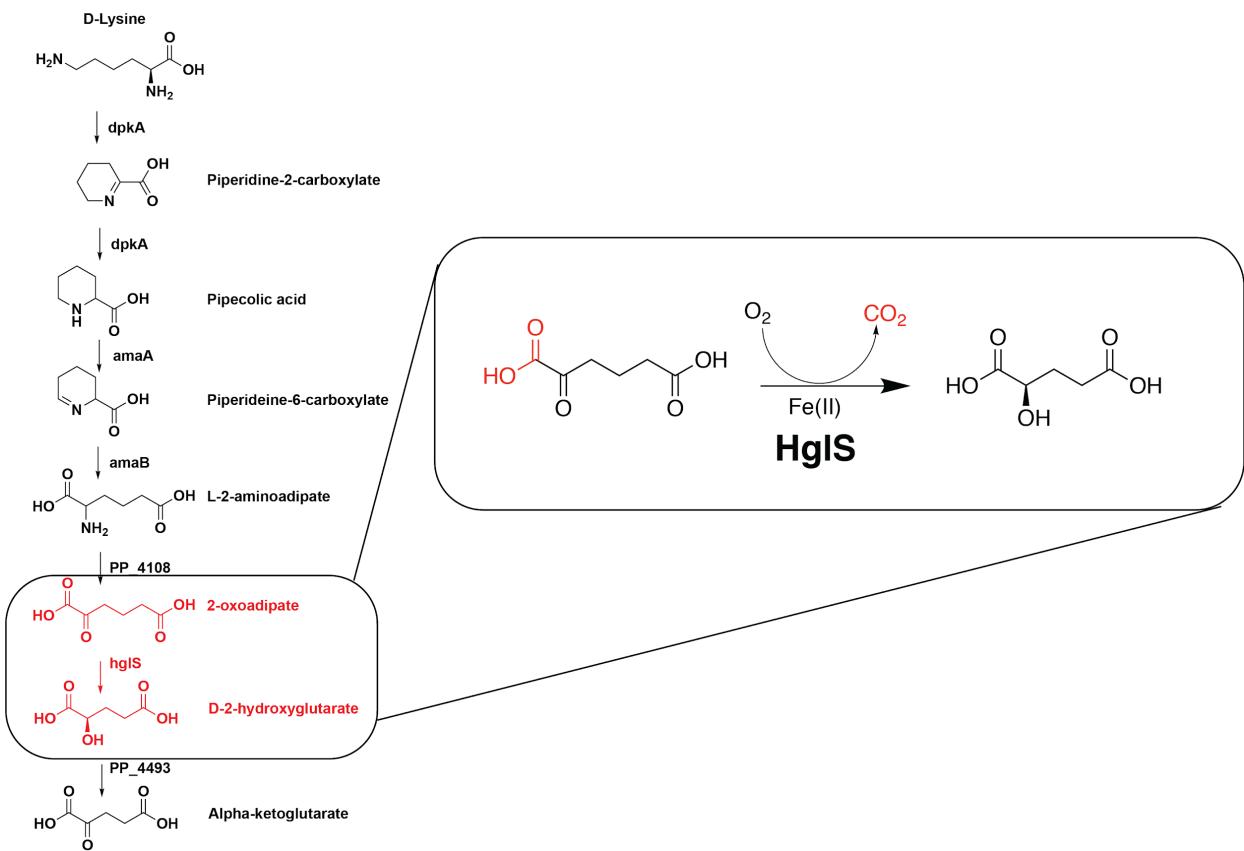
*Acad Sci USA* **87**, 4645–4649 (1990).

### **Competing Interests**

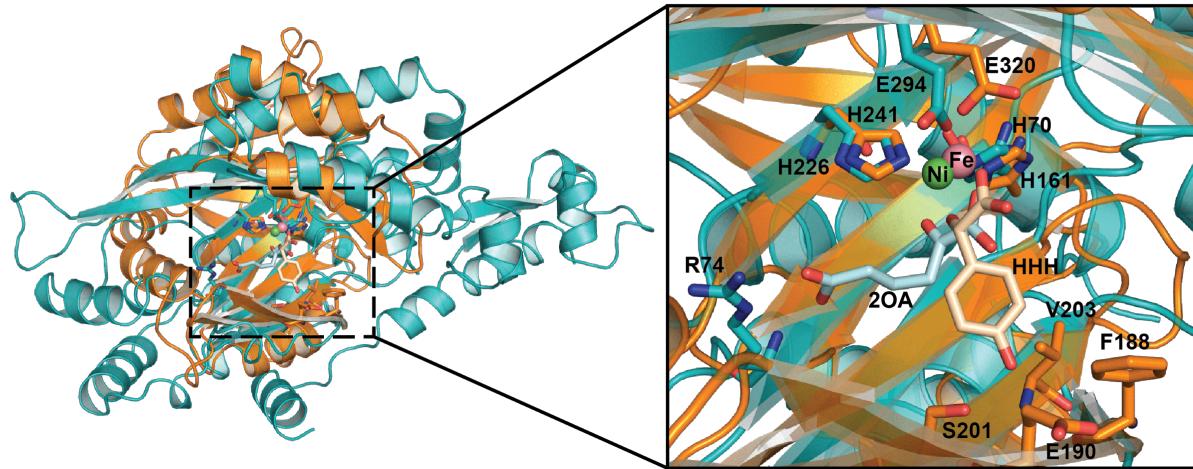
J.D.K. has financial interests in Amyris, Lygos, Demetrix, Napigen, Maple Bio, Apertor Labs and Ansa Biotechnology.

## Supporting Information

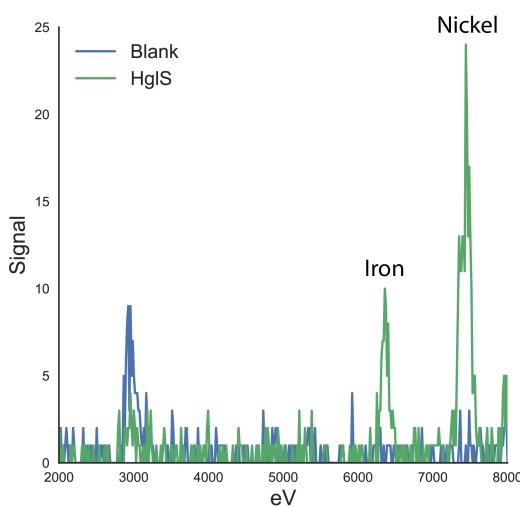
### Supporting Figures



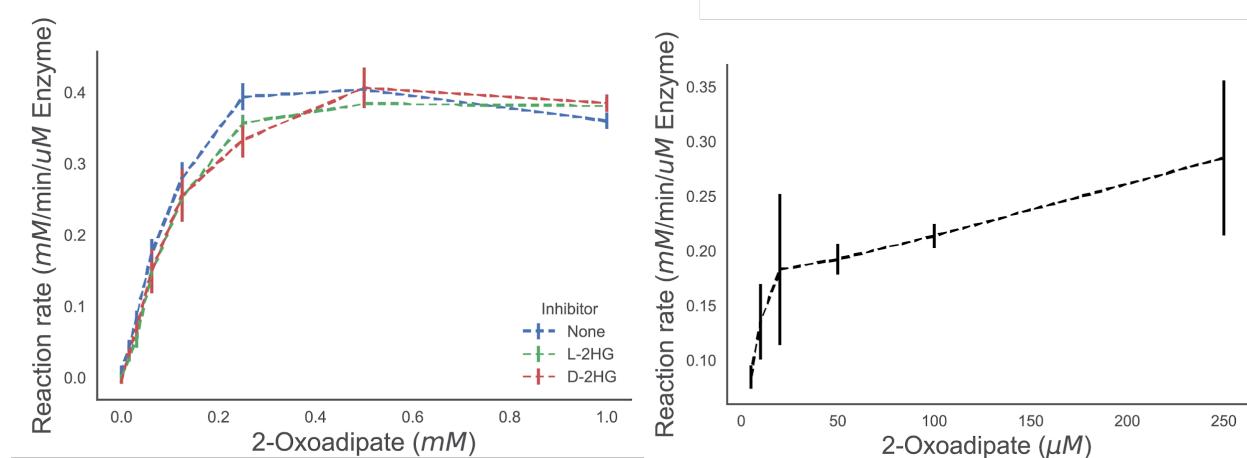
**Figure S1: D-lysine catabolic pathway of *Pseudomonas putida*.** D-lysine is catabolized independently of L-lysine in *Pseudomonas putida*. PP\_4108 transaminates 2AA to 2OA. HglS then catalyzes a previously undescribed 2OA to D-2HG decarboxylation (see inset).<sup>8</sup> D-2HG is then oxidized to alpha-ketoglutarate, connecting lysine metabolism to central metabolism (the TCA cycle).



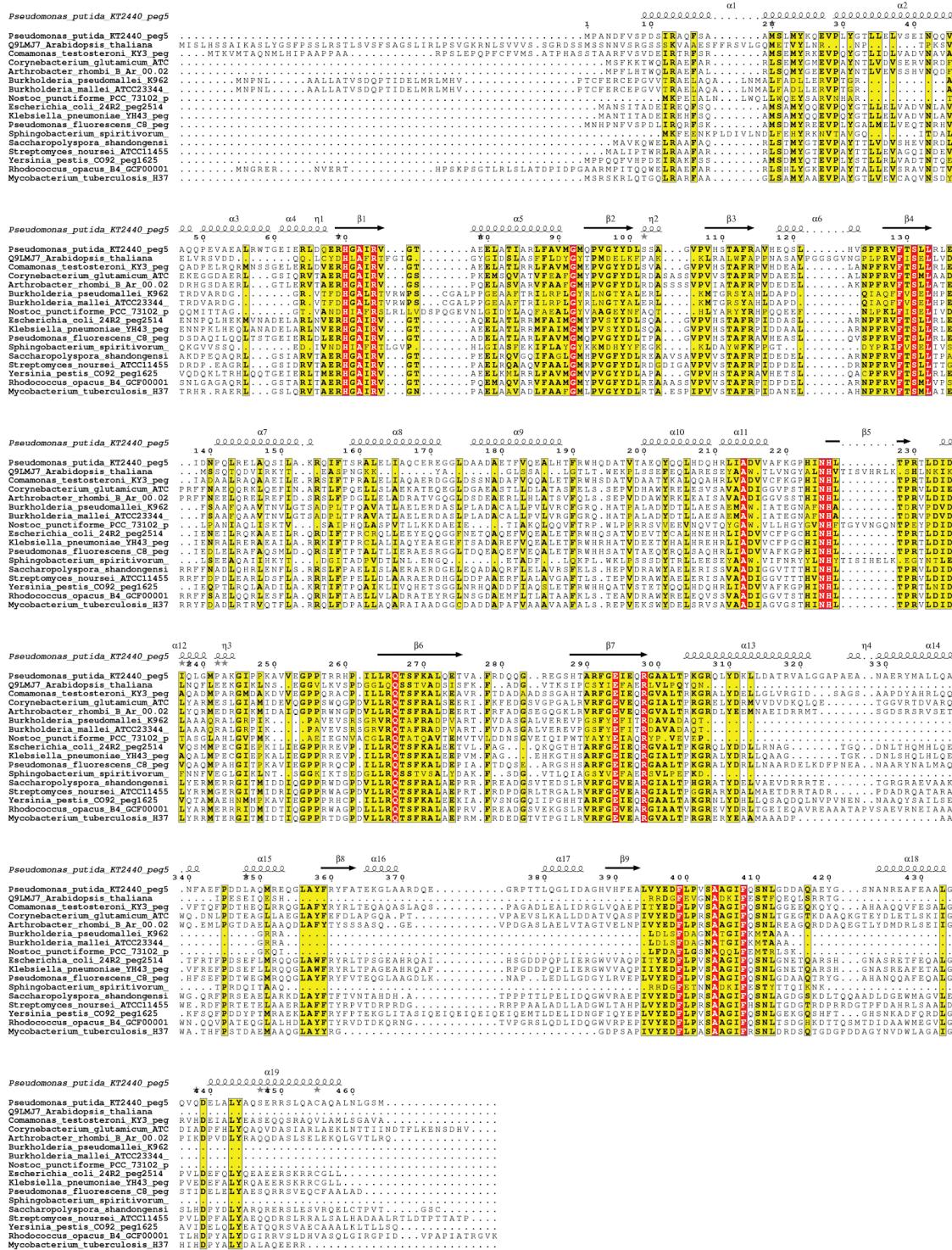
**Figure S2: Overlay of HgLS and HMS.** Overlay of the active sites of HgLS (teal) and HMS (orange) showing the conservation of the central  $\beta$ -sheet domain and active site architecture. The inset shows a magnified image of the active site in which metal-binding residues, substrates, and residues that interact with substrates in either enzyme are shown as sticks. The metal cofactors for HgLS (green) and HMS (pink) are labeled with the appropriate chemical symbol. The substrates are labeled as 2OA (2-oxoadipate) and HHH ((2S)-hydroxy(4-hydroxyphenyl)ethanoic acid) and shown in light blue and tan, respectively.



**Figure S3: K<sub>α</sub> emission spectrum of HgIS.** K-alpha emission spectrum resulting from a fluorescent scan of an HgIS crystal. The emitted wavelengths were characteristic of Ni (near 7475 eV), as well as Fe (near 6400 eV).



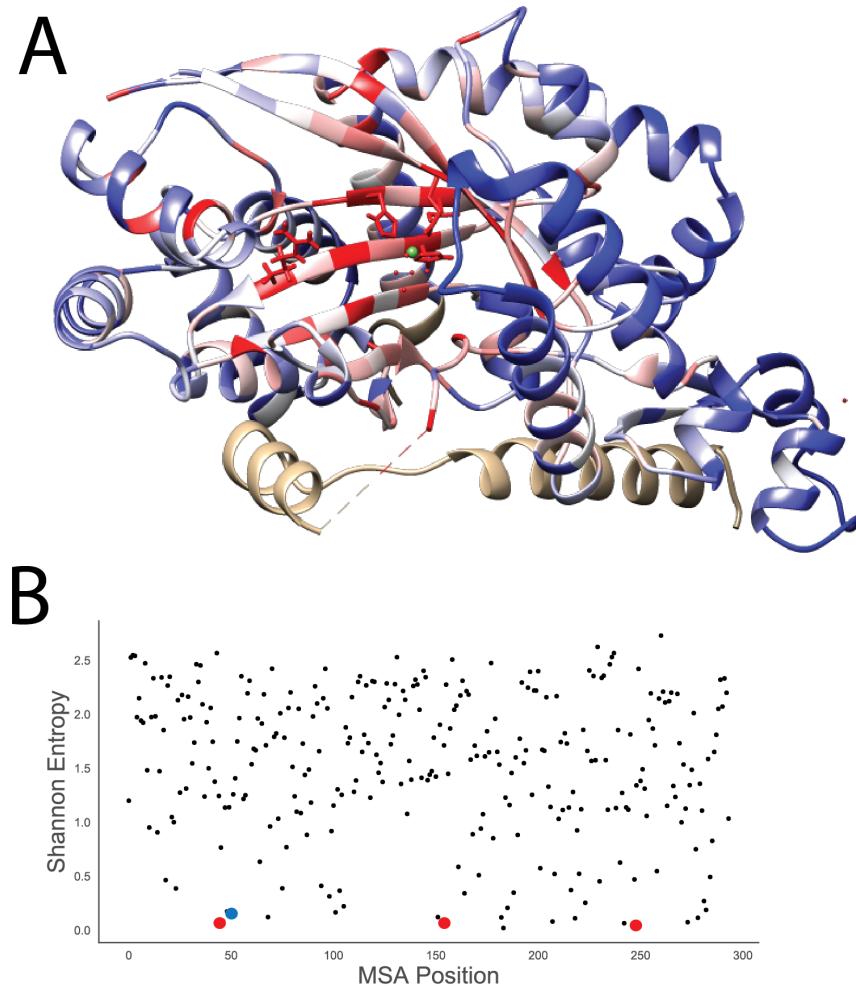
**Figure S4: Feedback inhibition and O<sub>2</sub> kinetics of HgIS.** **Left:** Reaction rates of HgIS on 2OA in the presence of either L-2HG (green) or D-2HG (red) compared to a control with no added product (blue) as measured by enzyme coupled decarboxylation assay, n=3. **Right:** Reaction rates of HgIS measured by oxygen consumption, n=3. The K<sub>m</sub> was determined to be 0.01 mM with a V<sub>max</sub> of 0.273 mM/min/μM enzyme.



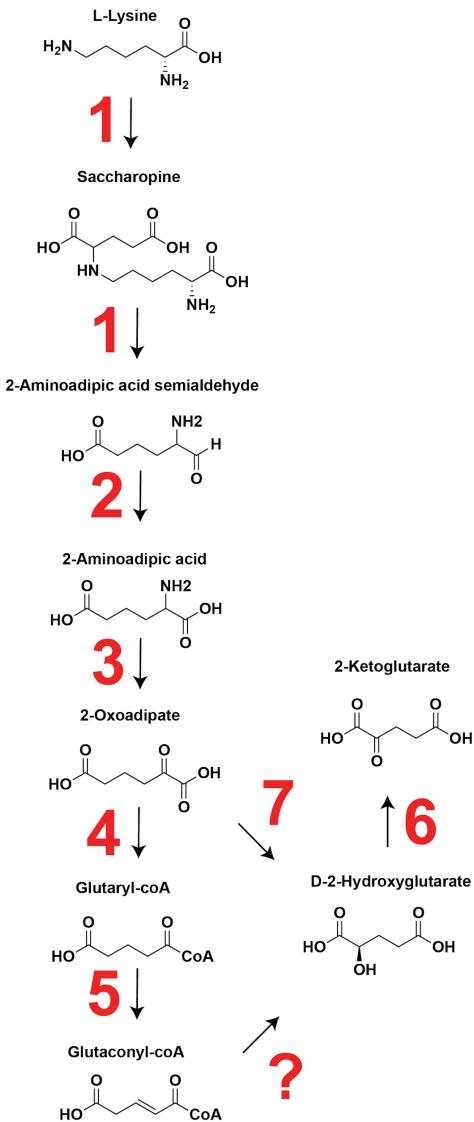
**Figure S5 - Alignment of bacterial and plant homologs of *P. putida* HglS (PP\_5260).**

Conserved residues are shown in yellow, while perfectly conserved residues are shown in red.

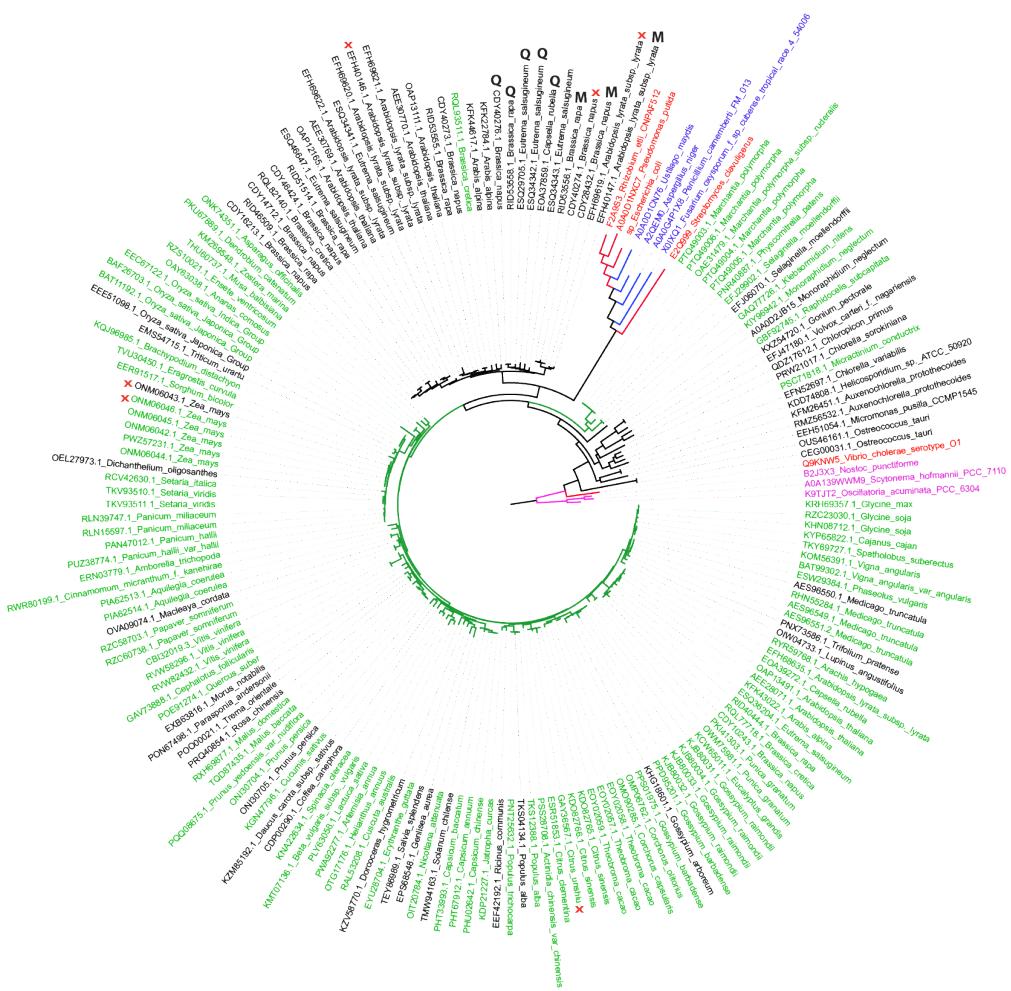
Above alignments is the secondary structure of HglS.



**Figure S6 - Conservation of critical residues in DUF1338 proteins.** A) Structure of HglS with residues colored by conservation, red showing the highest level of conservation with blue showing the least amount of conservation B) Shannon entropy of an alignment of the representative DUF1338 containing proteins in the Pfam database. The Fe (II) coordinating HHE triad residues are highlighted as red dots, while the carboxylate coordinating arginine is highlighted as a blue dot.

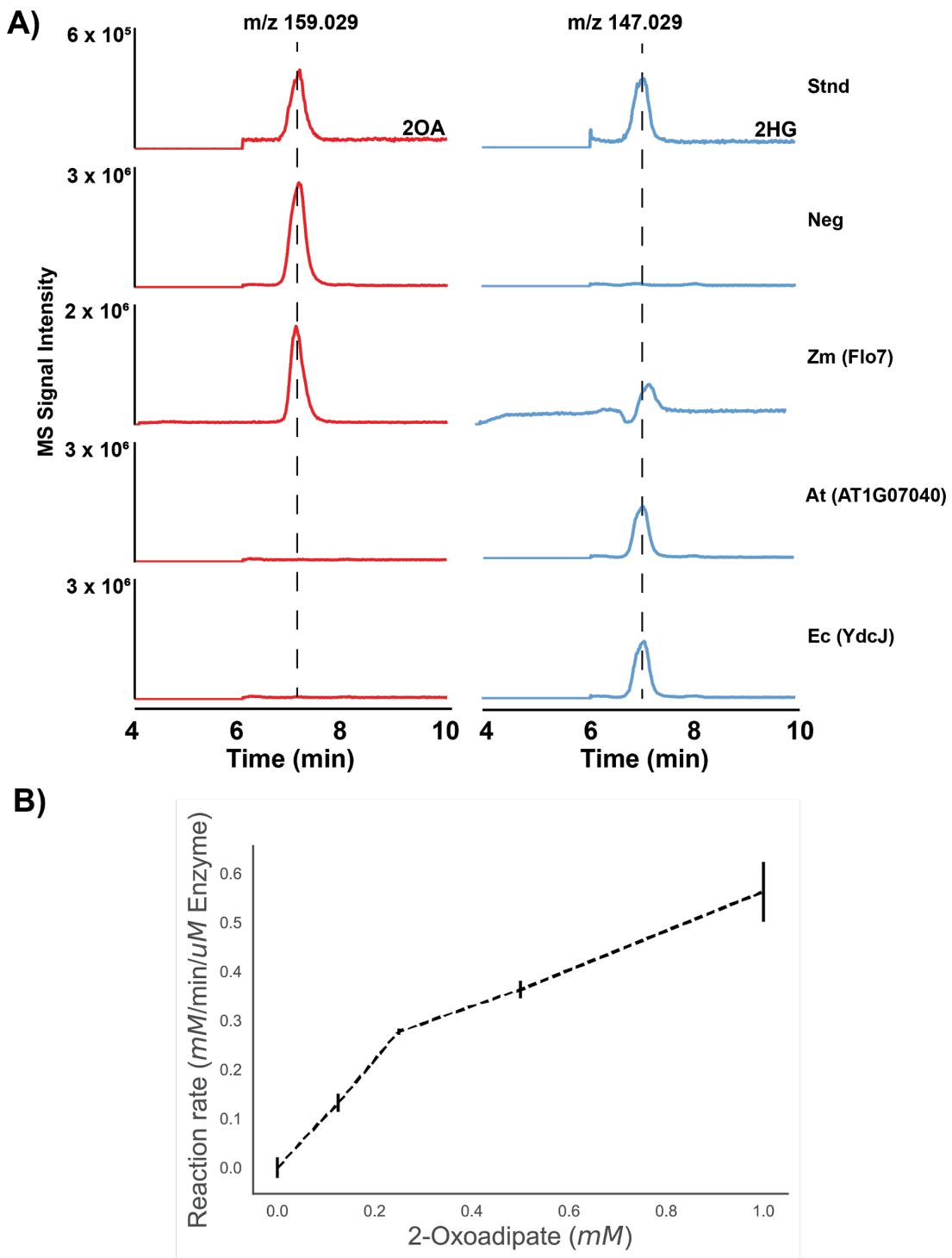


**Figure S7 - Putative route of lysine catabolism in plants.** Enzymes putatively involved in plant lysine catabolism according to <sup>7</sup> 1) Lysine-2KG reductase/saccharopine dehydrogenase 2) Aldehyde dehydrogenase 3) Gamma-aminobutyrate aminotransferase 4) 2KGD 5) Isovaleryl-coA dehydrogenase 6) D-2-hydroxyglutarate dehydrogenase. ? represents a putative enzyme that would convert glutaconyl-coA to D2HG. We propose that HglS homologs (7) catalyze the conversion of 2OA to 2HG in plants, rather than proceeding through the 2KGD complex.



**Figure S8 - Phylogenetic relationship of plant DUF1338 homologs.** Phylogenetic tree

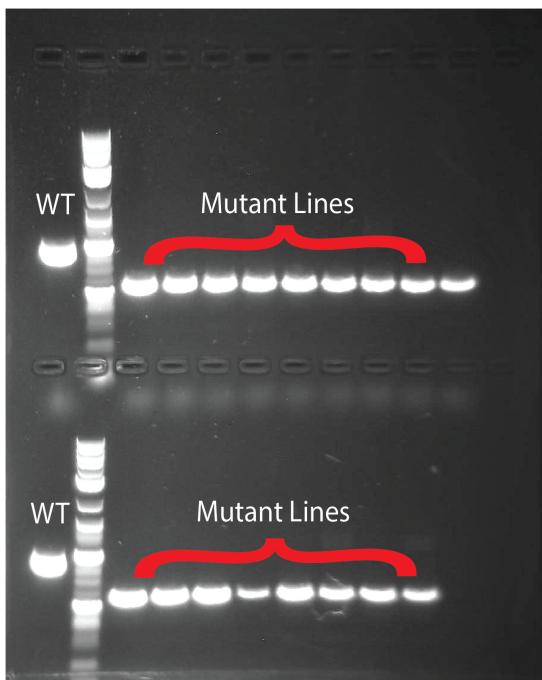
showing plant DUF1338 containing proteins in addition to bacterial (red), cyanobacterial (pink), and fungal (blue) homologs. Green leaves show homologs that either have a predicted mitochondria or chloroplast localization sequence, while black leaves do not have any predicted localization tag. Red X's show proteins without a conserved "HHE" triad, while M and Q show proteins that have substitutions at the predicted specificity residue.



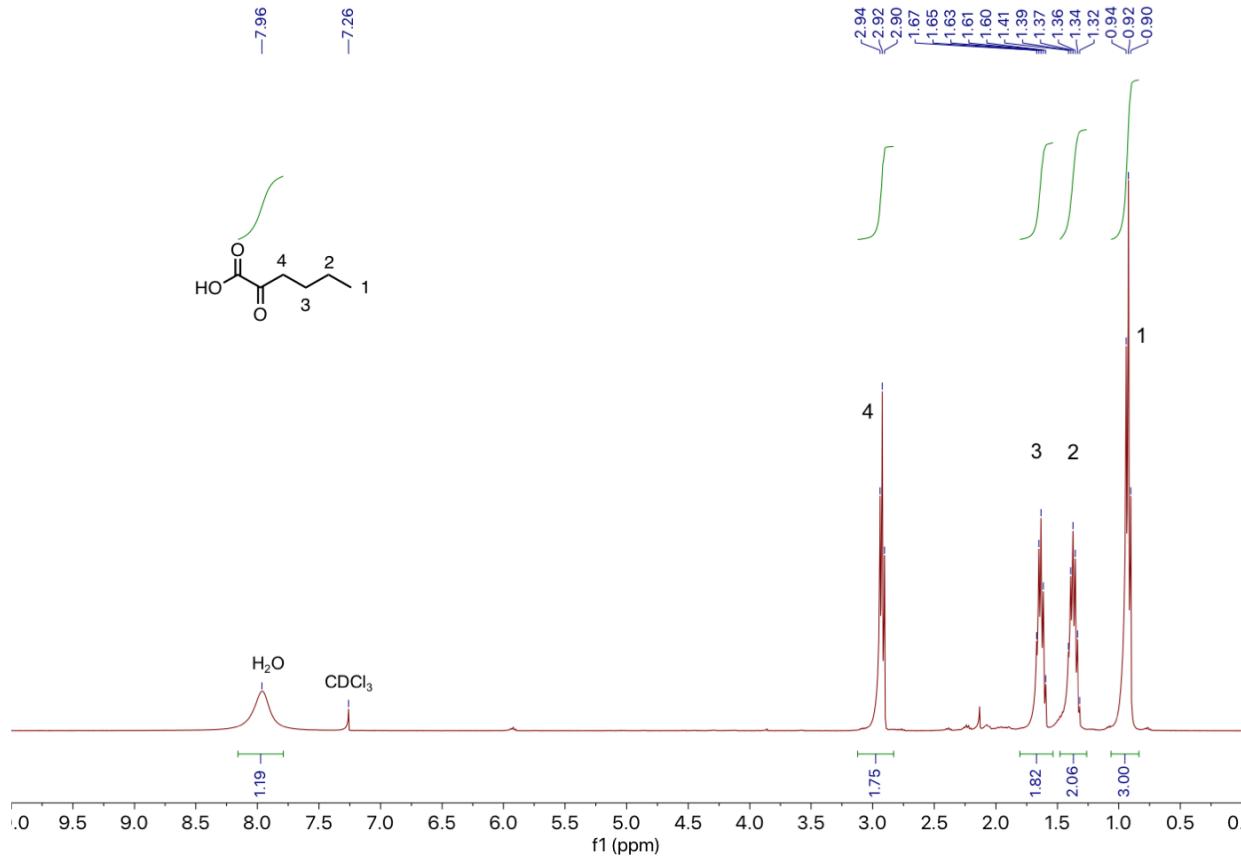
**Figure S9 - Biochemical analysis of plant HgLS homologs: A)** Extracted ion chromatograms (EICs) from high resolution LC-TOF analysis of biochemical assays of HgLS homologs performed in negative ion mode. EICs pertaining to 2-oxoadipate (2OA,  $m/z$  159.029, left

column) are shown in red, while EICs pertaining to 2-hydroxyglutarate (2HG, m/z 147.029, right column) are shown in blue. Assays were compared to authentic standards (Stnd) and negative control reactions (Neg) containing no enzyme. Data shown pertain to assays performed with the following homologs: *Zea mays*/Flo7 (Zm(Flo7)), *Arabidopsis thaliana*/AT1G07040 (At(AT1g07040)), and *Escherichia coli*/YdcJ (Ec(YdcJ)). The chromatograms shown are a representative sample from assays performed in triplicate. **B)** Flo7 enzyme turnover rates when provided with varying concentrations of 2OA substrate as measured by an enzyme coupled assay. Error bars represent 95% confidence interval, n = 3

**Figure S10: Confirmation of SALK lines -** PCR confirmation of homozygosity in Salk\_103299C mutant plants



**Figure S11: NMR of 2-oxohexanoic acid**



## Supplemental Tables

**Table S2 - Distribution of DUF1338 proteins across domains of life and conservation of key residues.**

	<b>DUF1338</b>	<b>HHE Triad</b>	<b>R74</b>
<b>Total</b>	2417	2070	2058
<b>Plant</b>	150	122	114

<b>Bacteria</b>	2074	1773	1769
<b>Fungi</b>	184	172	172

**Table S3 - Predicted localization of plant DUF1338 proteins in sequenced plant genomes**

	<b>Total</b>	<b>HHE</b>	<b>R</b>	<b>Chlor Localized Only</b>	<b>Mito Localized Only</b>	<b>Either Predictio n</b>	<b>No Signal Prediction</b>
Plants	202	167	159	61	8	117	85

**Table S4: Strains and plasmids used in this study**

<b>Strain</b>	<b>JBEI Part ID</b>	<b>Reference</b>
E. coli DH10B		<sup>67</sup>
E. coli BL21(DE3)		Novagen
<b>Plasmids</b>		
pET21b		Novagen
pET21b-PP_5260		<sup>8</sup>
pET21b-PP_5260 R74A	TBD	This work
pET21b-PP_5260 V402P	TBD	This work

pET21b- <i>FLO7</i>	TBD	This work
pET21b- <i>FLO7</i> $\Delta$ 1-76	TBD	This work
pET21b-AT1G07040 $\Delta$ 1-56	TBD	This work
pET21b- <i>ydcJ</i> _ECO	TBD	This work

**Table S5: Summary of X-ray data collection statistics HgIS apo-enzyme, HgIS substrate-bound and Flo7 substrate-bound.**

	<b>HgIS apo</b>	<b>HgIS substrate-bound</b>	<b>Flo7 substrate-bound</b>
<b>Wavelength</b>	1.00000	0.99999	0.97946
<b>Resolution range</b>	50.37 - 1.14 (1.181 - 1.14)	72.05 - 1.419 (1.47 - 1.419)	38.65 - 1.85 (1.916 - 1.85)
<b>Space group</b>	P 21 21 21	P 21 21 21	P 21 21 21
<b>Unit cell</b>	42.44 100.74 104.76 90 90 90	42.44 100.28 103.57 90 90 90	92.30 98.20 203.80 90 90 90
<b>Total reflections</b>	1121754 (103068)	593341 (56327)	1141314 (116637)
<b>Unique reflections</b>	164034 (16167)	84333 (8301)	156493 (15364)
<b>Multiplicity</b>	6.8 (6.4)	7.0 (6.8)	7.3 (7.6)
<b>Completeness (%)</b>	99.94 (99.40)	99.84 (99.10)	98.94 (98.68)
<b>Mean I/sigma(I)</b>	12.33 (2.10)	16.37 (1.36)	11.52 (1.39)
<b>Wilson B-factor</b>	13	15.71	29.19
<b>R-merge</b>	0.0615 (0.6688)	0.06148 (1.42)	0.1162 (1.592)
<b>R-meas</b>	0.06648 (0.7282)	0.06641 (1.538)	0.1251 (1.707)
<b>R-pim</b>	0.0249 (0.2833)	0.02487 (0.5859)	0.04555 (0.6084)
<b>CC1/2</b>	0.999 (0.666)	0.999 (0.736)	0.998 (0.517)
<b>CC*</b>	1 (0.894)	1 (0.921)	1 (0.825)
<b>Reflections used in refinement</b>	164033 (16166)	84229 (8250)	156474 (15364)
<b>Reflections used for R-free</b>	8263 (802)	4041 (367)	7752 (777)
<b>R-work</b>	0.1499 (0.2438)	0.1690 (0.2796)	0.1711 (0.3159)

<b>R-free</b>	0.1736 (0.2776)	0.1990 (0.3251)	0.1949 (0.3405)
<b>CC(work)</b>	0.967 (0.856)	0.972 (0.829)	0.961 (0.771)
<b>CC(free)</b>	0.962 (0.826)	0.955 (0.739)	0.958 (0.757)
<b>Number of non-hydrogen atoms</b>	4198	4264	11051
<b>Macromolecules</b>	3461	3612	9800
<b>Ligands</b>	1	2	24
<b>Solvent</b>	736	640	1123
<b>Protein residues</b>	434	449	1236
<b>RMS(bonds)</b>	0.009	0.008	0.009
<b>RMS(angles)</b>	1.07	1.13	1.15
<b>Ramachandran favored (%)</b>	98.36	97.74	98.03
<b>Ramachandran allowed (%)</b>	1.64	2.26	1.89
<b>Ramachandran outliers (%)</b>	0	0.00	0.08
<b>Rotamer outliers (%)</b>	1.67	1.33	0
<b>Clashscore</b>	3.91	7.33	3.54
<b>Average B-factor</b>	23.03	26.84	37.70
<b>Macromolecules</b>	20.91	25.45	36.15
<b>Ligands</b>	13.40	21.62	89.45
<b>Solvent</b>	33.05	34.77	45.40